

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 07:19:15 ; Search time 16.6794 seconds
(without alignments)
234.715 Million cell updates/sec

Title: US-09-787-082-8

Perfect score: 119

Sequence: 1 GCCSNPVCHLEHSLNCTNG 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	63	52.9	694	5 Q95XU6	Q95xu6 caenorhabdi
2	57.5	48.3	504	5 Q9XZX8	Q9xxz8 leishmania
3	57.5	48.3	709	5 Q9XZX9	Q9xxz9 leishmania
4	53	44.5	61	5 Q9BP56	Q9bp56 conus penna
5	50.5	42.4	966	5 Q22378	Q22378 caenorhabdi
6	50	42.0	61	5 Q9BP57	Q9bp57 conus penna
7	50	42.0	2104	5 Q21281	Q21281 caenorhabdi
8	50	42.0	2104	5 Q964N4	Q964n4 caenorhabdi
9	49.5	41.6	151	5 Q17681	Q17681 caenorhabdi
10	49	41.2	451	13 Q9PT49	Q9pt49 atractaspis
11	49	41.2	555	4 Q96K49	Q96k49 homo sapien
12	49	41.2	569	11 Q88716	Q88716 mus musculus
13	49	41.2	751	11 Q9R218	Q9r218 mus musculus
14	49	41.2	751	11 Q9QYB2	Q9qyb2 mus musculus
15	48.5	40.8	130	16 Q9KP39	Q9kpb39 vibrio chol
16	48.5	40.8	136	11 Q99PK1	Q99pk1 mus musculus

17	48.5	40.8	643	11 Q9ERV6	Q9erv6 mus musculus
18	48.5	40.8	655	11 Q9WVF5	Q9wvf5 mus musculus
19	48.5	40.8	1209	11 Q9XQ70	Q9qx70 rattus norv
20	48.5	40.8	1210	11 Q9EP98	Q9ep98 mus musculus
21	48	40.3	160	5 Q9XU68	Q9xu68 caenorhabdi
22	48	40.3	199	10 Q49095	Q49095 leavenworth
23	48	40.3	199	10 Q49096	Q49096 leavenworth
24	48	40.3	199	10 Q49097	Q49097 leavenworth
25	48	40.3	199	10 Q49098	Q49098 leavenworth
26	48	40.3	199	10 Q49099	Q49099 leavenworth
27	48	40.3	199	10 Q49100	Q49100 leavenworth
28	48	40.3	199	10 Q49101	Q49101 leavenworth
29	48	40.3	199	10 Q49108	Q49108 leavenworth
30	48	40.3	199	10 Q50045	Q50045 leavenworth
31	48	40.3	223	10 Q49111	Q49111 leavenworth
32	48	40.3	226	10 Q9LD74	Q9ld74 arabidopsis
33	48	40.3	238	10 Q9M4B3	Q9m4b3 arabidopsis
34	48	40.3	281	10 Q49113	Q49113 leavenworth
35	48	40.3	322	10 Q49115	Q49115 leavenworth
36	48	40.3	322	10 Q49114	Q49114 leavenworth
37	48	40.3	338	10 Q49109	Q49109 leavenworth
38	48	40.3	347	5 Q8SYW5	Q8syw5 drosophila
39	48	40.3	350	10 Q9ZWK4	Q9zwk4 brassica ol
40	48	40.3	357	10 Q9M4B2	Q9m4b2 arabidopsis
41	48	40.3	359	10 Q9ZWLO	Q9zwlo arabis hirs
42	48	40.3	361	10 Q9LDM3	Q9ldm3 arabidopsis
43	48	40.3	361	10 Q9ZWL4	Q9zwl4 arabis stel
44	48	40.3	361	10 Q9ZWL2	Q9zwl2 arabis flag
45	48	40.3	361	10 Q9ZWL1	Q9zwl1 arabidopsis

ALIGNMENTS

RESULT 1

Q95XU6 ID Q95XU6 PRELIMINARY; PRT; 694 AA.

AC Q95XU6;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Hypothetical 76.9 kDa protein.

GN Y61A9LA.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.,"

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Ali J., Dempsey S.;

RT "The sequence of C. elegans cosmid Y61A9LA.,"

RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RT "Direct Submission.,"

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC024843; AAK70666.2; -

KW Hypothetical protein.

SQ SEQUENCE 694 AA; 76927 MW; BF1A5C0CC2DB0FDE CRC64;

Query Match 52.9%; Score 63; DB 5; Length 694;

Best Local Similarity 47.4%; Pred. No. 0.047;

```
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCCSNPVCHEHSLNCTNG 19
    | | | | | | | | | |
Db 632 GACRNPICFKRHKPCREG 650

RESULT 2
Q9XZX8 PRELIMINARY; PRT; 504 AA.
AC Q9XZX8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Possible surface antigen.
GN L302.01.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
CC -!- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL; AL389894; CAC22669.1; -.
DR InterPro; IPR000564; 2FE2S_ferredoxin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002049; Laminin_EGF.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW Iron-sulfur.
SQ SEQUENCE 504 AA; 52823 MW; B75450E5FE599263 CRC64;

Query Match 48.3%; Score 57.5; DB 5; Length 504;
Best Local Similarity 50.0%; Pred. No. 0.28;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

QY 3 CSNPV---CHLEHSLNCTNG 19
    | | | | | | | | | |
Db 398 CSPTTQPCVEHCNTCVNG 417

RESULT 3
Q9XZX9 PRELIMINARY; PRT; 709 AA.
AC Q9XZX9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Possible surface antigen.
GN L302.02.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Rajandream M.A., Barrell B.G.;
```

```
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL389894; CAC22670.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00261; FU; 4.
SQ SEQUENCE 709 AA; 73961 MW; C6BIA5348B9B1637 CRC64;

Query Match 48.3%; Score 57.5; DB 5; Length 709;
Best Local Similarity 50.0%; Pred. No. 0.38;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

QY 3 CSNPV---CHLEHSLNCTNG 19
    | | | | | | | | | |
Db 603 CSPTTQPCVEHCNTCVNG 622

RESULT 4
Q9BP56 PRELIMINARY; PRT; 61 AA.
AC Q9BP56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Conotoxin scaffold 1.
OS Conus pennaceus (Feathered cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=37335;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21105969; PubMed=11158371;
RA Conticello S.G., Gilad Y., Avidan N., Ben-Asher E., Levy Z.,
RA Fainzilber M.;
RT "Mechanisms for Evolving Hypervariability: The Case of Conopeptides.";
RL Mol. Biol. Evol. 18:120-131(2001).
DR EMBL; AF215089; AAG60510.1; -.
DR HSP; P50985; 1AKG.
SQ SEQUENCE 61 AA; 6348 MW; 6FCFD333D3D306921 CRC64;

Query Match 44.5%; Score 53; DB 5; Length 61;
Best Local Similarity 50.0%; Pred. No. 0.24;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCCSNPVCHEHSLNLC 16
    | | | | | | | | | |
Db 45 GCCSRPPCALSNPDYC 60

RESULT 5
Q22378 PRELIMINARY; PRT; 966 AA.
AC Q22378;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 102.5 kDa protein.
GN T10E10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
```

```

RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C.;
RT "The sequence of C. elegans cosmid t10E10.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39644; AAA80360.2; -.
DR HSSP; P10969; IWGT.
DR InterPro; IPR002557; Chitin_bind_Pera.
DR InterPro; IPR007094; Ketoacyl-synt.
DR InterPro; IPR003571; Snake toxin.
DR InterPro; IPR002899; Wrl/EB.
DR Pfam; PF01607; CBM_14; 2.
DR SMART; SM00289; Wrl; 12.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
DR PROSITE; PS00272; SNAKE_TOXIN; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 966 AA; 102460 MW; B565A3CDD25216D9 CRC64;

Query Match 42.4%; Score 50.5; DB 5; Length 966;
Best Local Similarity 33.3%; Pred. No. 6.8;
Matches 10; Conservative 3; Mismatches 6; Indels 11; Gaps 1;

Qy 1 GCC-----SNPVCHLEHSLNCTNG 19
||| ||| ||| :|| :|
Db 705 GCCLLSMFVCPTRSNVAVCQSPNNVCPSG 734

RESULT 6
Q9BP57 PRELIMINARY; PRT; 61 AA.
AC Q9BP57;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Conotoxin scaffold I.
OS Conus pennaceus (Feathered cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=37335;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21105969; PubMed-11158371;
RA Conticello S.G., Gilad Y., Avidan N., Ben-Asher E., Levy Z.,
RA Fainzilber M.;
RT "Mechanisms for Evolving Hypervariability: The Case of Conopeptides.";
RL Mol. Biol. Evol. 18:120-131(2001).
DR EMBL; AF215088; AAG60509.1; -.
DR HSSP; P50985; IAKG.
SQ SEQUENCE 61 AA; 6363 MW; 42E003324D66922 CRC64;

Query Match 42.0%; Score 50; DB 5; Length 61;
Best Local Similarity 50.0%; Pred. No. 0.73;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GCCSNPVCHLEHSLNLC 16
||| | | | | |
Db 45 GCCSLPPCALSNPDYC 60

RESULT 7
Q21281 PRELIMINARY; PRT; 2104 AA.
AC Q21281;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 230.1 kDa protein.
GN K07D8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello T.;
RT "The sequence of C. elegans cosmid K07D8.";
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; L16679; AAA28092.5; -.
DR HSSP; P02468; 1TLE.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000361; EGF-like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR000082; SEA_domain.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF; 20.
DR Pfam; PF01390; SEA; 4.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00181; EGF; 35.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00200; SEA; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_15.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS50024; SEA; 4.
DR PROSITE; PS50234; VWF_A; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hypothetical protein;
KW Repeat.
SQ SEQUENCE 2104 AA; 230051 MW; C98CD462CC434F8B CRC64;

Query Match 42.0%; Score 50; DB 5; Length 2104;
Best Local Similarity 41.2%; Pred. No. 16;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 3 CSNPVCHLEHSLNCTNG 19
| | | | | | | |
Db 1286 CGNGICHLDLGEVCGG 1302

RESULT 8
Q964N4 PRELIMINARY; PRT; 2104 AA.
AC Q964N4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Transmembrane matrix receptor MUP-4.
```


DR EMBL; AE004323; AAF95681.1; -.
 DR TIGR; VC2540; -.
 DR InterPro; IPR002221; WAP.
 DR PROSITE; PS00317; 4_DISULFIDE_CORE; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 130 AA; 14262 MW; 1A56F795EE007A86 CRC64;

Query Match 40.8%; Score 48.5; DB 16; Length 130;
 Best Local Similarity 61.1%; Pred. No. 2.5;
 Matches 11; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 GCCSNPVCCHLEHSLCTN 18
 |||||
 Db 95 GCVSNPMCL-CSVCTN 111

Search completed: March 17, 2003, 07:26:24
 Job time : 18.6794 secs